一种基于胶囊网络的图像检索方法

黄 静　杨树国*　刘子正

（青岛科技大学数理学院，山东　青岛　266061）

摘　要：作为一种新兴的网络结构，胶囊网络用向量输出替代标量输出，能够捕捉图像特征之间的空间关系，改善卷积神经网络的局限性。首先对胶囊网络进行训练实现图像分类，得到图像的预测类标签，判定出查询图像的所属类别，然后将网络的数字胶囊层中的特征参数作为图像的特征向量，在查询图像的所属类别集合中利用图像特征向量找到与查询图像相似的图像。分别在FASHION-MNIST和CIFAR10数据集上进行了实验，实验结果表明本文方法可以较好地提取出图像的特征，分别提升了查准率，并取得了良好的图像检索结果。

关键词：图像检索；胶囊网络；特征提取

中图分类号：TP391.41　　**文献标识码：**A　　**文章编号：**1008 - 6609 (2020) 06 - 0014 - 05

DOI:10.15966/j.cnki.dnydx.2020.06.004

1　引言

近几年，神经网络的相关研究见证了基于学习的特征在图像处理、音频识别和自然语言处理等多个领域的成功，其中，基于内容的图像检索(Content-Based Image Retrieval，CBIR)研究更是借助于神经网络和深度学习等方法实现了蓬勃发展[1,2]，并且其相关技术可直接应用于各种计算机视觉任务中，如人脸识别[3]、目标检测[4]和语义分割[5]等，使得CBIR具有重要的现实研究意义。卷积神经网络(Convolutional Neural Network，CNN)作为神经网络在图像处理领域中应用的典型代表，利用网络学习获得接近人类认知的高层特征，明显改善了图像分类以及目标检测等领域的实验效果。

尽管CNN已经战胜了许多基于图像特征的工艺性设计方法，如SIFT、HOG、灰度共生矩阵，但仍然存在一些不足之处。CNN通常包含多个网络层，每一个网络层中都会产生一系列的特征，从边缘、区域到整个实例对象，逐渐学习图像的特征。但是，CNN对于图像中的仿射变换学习能力较差，没有考虑图像内部的空间关系，并且由于池化层的存在，丢失了图像的局部特征信息。因此，为了提高卷积网络的泛化能力，通常采用图像扩增的方式增大训练集的规模，来提高网络的训练精度，但这种方法将消耗更多的计算时间和存储资源。此外，CNN还容易受到白盒攻击[6]，例如快速梯度符号攻击[7]。

为了消除CNN的局限性，Hinton等人[8]提出了一种新的网络结构——胶囊网络(Capsule Network，CapsNet)，采用胶囊层之间的动态路由来取代池化层在网络中的作用。CapsNet是由一组胶囊形成的网络结构：一个胶囊是一组神经元，每个胶囊负责识别实体的特定属性，胶囊中激活的神经元决定了图像的特征。考虑到CNN的基本思想是学习特征检测器的平移复制，换句话说，在一个位置收集到的经过训练的特征信息可以扩展到其他位置，并且这一功能有利于收集图像的底层信息。而CapsNet不仅继承了CNN检测性学习的优点，还用向量输出替代了标量输出，将图像的部分区域作为输入，检测实例化对象的存在和姿态，利用路由协议取代子采样(最大池化或平均池化)，较好地保存了图像的局部空间结构。胶囊一个突出的特点是，在图像特征表达方面以等变性取代不变性，这使网络在没有数据扩增的情况下对仿射变换具有良好的学习能力。CapsNet不仅在MNIST数据集上取得了超过CNN方法的有史以来最高的分类准确率，并且在重叠数字识别方面表现出了巨大的潜力。

本文的其余部分安排如下：第2节介绍Capsule的详细信息以及Routing算法的基本思想；第3节描述本文方法中所采

作者简介：黄静(1993-)，女，山东枣庄人，硕士，研究方向：图像检索。

***通讯作者：**杨树国(1970-)，男，山东曹县人，博士，教授，研究方向为图像检索、图像识别、语音识别等。

基金项目：2019年青岛科技大学大学生创新创业训练计划项目，项目名称：基于胶囊网络和卷积神经网络的目标识别方法研究，项目编号：X201910426241。

用的网络结构及图像检索算法；第4节进行基于CapsNet的图像检索实验，并对实验结果进行比较分析；第5节总结本文工作。

2 相关工作

2.1 胶囊单元

胶囊(Capsule)是一组神经元，其输入输出向量表示特定实体的实例化参数(特定物体、概念实体等出现的概率与某些属性)。同一层级的Capsule通过变换矩阵对更高级别的Capsule的实例化参数进行预测，当多个预测一致时，更高级别的Capsule将变得活跃。其中，向量的方向表示图像中存在的特定实体的各种性质，即实体的某些图形属性，这些性质可以包含多种不同的参数，例如姿势(方向，位置，大小)、变形、速度、色彩、纹理。向量的长度表示某个实体存在的概率，为了实现这种一致性预测，完成Capsule层级的激活功能，Hinton等人使用了一个被称为Squashing的非线性压缩函数，如公式(1)所示[8]。

$$v_j = \frac{\|s_j\|^2 s_j}{1+\|s_j\|^2\|s_j\|} \tag{1}$$

其中v_j为Capsule j的输出向量，s_j为上一层所有Capsule输入到当前层中Capsule$j$的向量加权和。该非线性函数可以看作是对向量长度的一种压缩，是一种输入向量激活后得到输出向量的方式。

如上所述，Capsule j的输入向量就相当于人工神经网络中神经元的标量输入，而该公式的计算就相当于两层Capsule间的传播与连接方式。输入向量s_j的计算过程可以用公式(2)表示。

$$s_j = \sum_i c_{ij}\hat{u}_{j|i} \qquad \hat{u}_{j|i} = W_{ij}u_i \tag{2}$$

其中u_i为上一层Capsule$i$的输出向量，W_{ij}为权重矩阵，c_{ij}为耦合系数，$\hat{u}_{j|i}$表示上一层Capsule$i$的输出向量u_i和对应的权重矩阵W_{ij}相乘而得出的预测向量。

在确定$\hat{u}_{j|i}$后，我们需要使用动态路由(Dynamic Routing)来计算输出向量s_j，然后将s_j投入到Squashing非线性函数得出Capsule$j$的输出向量v_j，其中，c_{ij}通过Routing算法迭代更新。

2.2 Dynamic Routing算法

在Capsule间加入一致性Routing机制可以找到一组系数c_{ij}，使得网络在学习过程中找到与输出向量v_j一致性最高的输入向量$\hat{u}_{j|i}$。c_{ij}通过Softmax函数确定，如公式(3)所示，并且上一层中某一Capsule$i$和后一层所有Capsule$j$之间的耦合系数和为1。

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{3}$$

这里b_{ij}是Capsule$i$与Capsule$j$耦合的对数先验概率，其初始值为0，通过b_{ij}的值加上$\hat{u}_{j|i}$和v_j的标量积来更新b_{ij}的值。

Routing是首先通过公式(2)和公式(1)得出Capsule$j$的初始输出向量v_j，然后利用公式(3)更新c_{ij}，进一步修正向量s_j，得到新的向量v_j，然后重复更新c_{ij}，不断提高后Capsule内部的一致性。Routing使得在胶囊内部不需要使用反向传播而是直接计算输入向量与输出向量间的一致性更新参数，避免了网络训练过程中出现的梯度消失问题。

3 本文方法

3.1 本文所采用的网络模型

CapsNet由卷积层(Conv)，主胶囊层(PrimaryCaps)和数字胶囊层(DigitCaps)三部分组成，本文将以一个大小为28×28的图像为例对CapsNet的基本结构进行详细介绍。其中，Conv是标准的卷积层，利用256个9×9卷积核来获取图像的底层卷积特征，在步幅为1且没有填充的情况下，空间尺寸减小为20×20。Conv层将像素信息转换为局部区域的活动，然后将其输出向量输入到PrimaryCaps层中，后者是一个支持胶囊向量的改进卷积层。PrimaryCaps层使用步幅为2的9×9卷积核将空间尺寸从20×20减小为6×6，生成32个8维卷积胶囊。然后将其输出向量输入到DigitCaps层，该层采用转换矩阵将每个类的8维胶囊向量转换为16维胶囊向量。最终，DigitCaps层具有针对每个数字类别的16维向量。为了更好地提取图像的特征，本文实验中所采用的具体网络结构及参数如图1所示。

3.2 损失函数

为了允许数据集中多个类别存在，本文采用的损失函数如公式(4)和(5)所示。其中，L_k是Capsule$k$的边际损失函数，在DigitCaps层中若第k类数字胶囊v^k存在则$T_k=1$，否则$T_k=0$。这里$m^+=0.9$，m^-用来控制损失函数的边界值，即如果第k类胶囊v^k存在，则存在的概率应该不小于0.9，否则存在的概率应该不超过0.1，$\lambda=0.5$。

总边际损失函数(*MarginLoss*)是DigitCaps层中所有数字胶囊的边际损失的总和。第二部分是重构损失函数(*ReconstructionLoss*)，它是重构图像和输入图像之间的平方差，重构损失使得网络保留重构图像所需的所有信息，在输入新的图像时，它还充当“调节器”，有助于防止网络过拟合。本文采用二维反卷积函数作为重构损失函数，并且为了防止反卷积过程中信息的过度丢失，步长设置不超过2，$\alpha=0.0005$用来连接边际损失函数和重构损失函数。

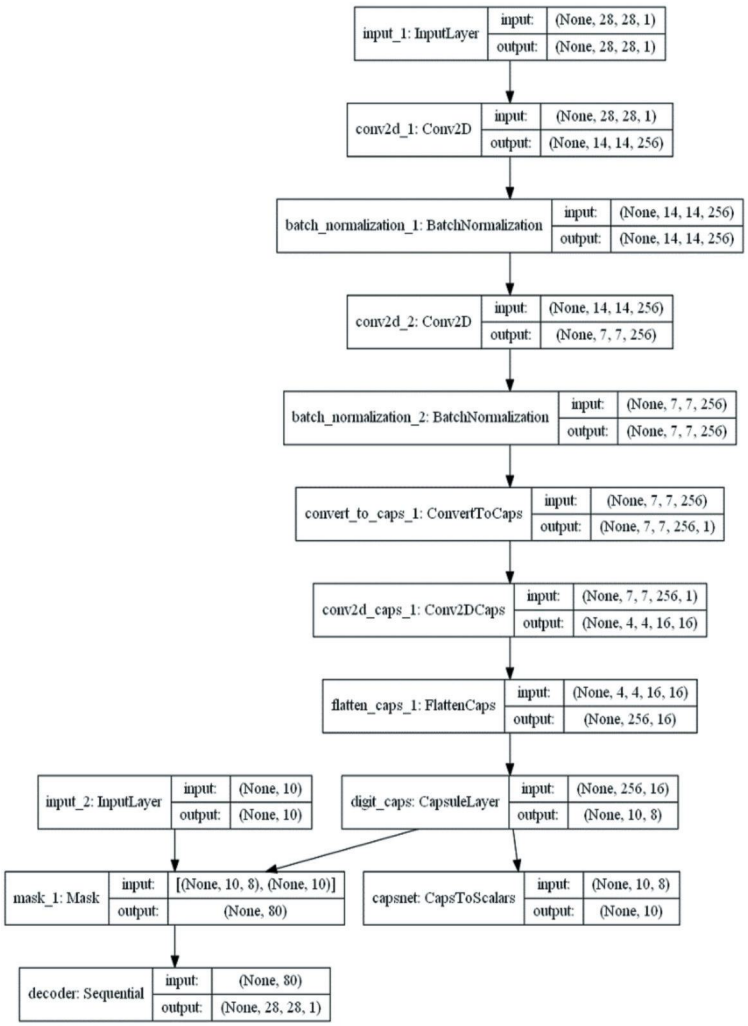


图1　本文所采用的网络结构

$$L_k = T_k \max(o, m^+ - \|v^k\|)^2 + \lambda(1 - T_k)\max(o, \|v^k\| - m^-)^2 \quad (4)$$

$$TotalLoss = MarginLoss + \alpha ReconstructionLoss \quad (5)$$

3.3　本文所采用的图像检索算法

本节将对本文所采用的基于CapsNet的图像检索算法的基本思想进行详细介绍。在图像检索实验中，训练集作为基本图像库，测试集作为样例图像库。首先，对上文中提到的网络模型进行参数训练，记录基本图像库与样例图像库中每一图像在最后一个训练周期中网络的DigitCaps层的特征参数，以及对应的预测类标签。然后，根据预测类标签对基本图像库中的图像进行分类，得到各个类别的子图像库，再将样例图像库中的样例图像划分到某一类别的子图像库中。最后，我们在该子图像库中查找与样例图像相似的图像。在检索实验过程中DigitCaps层中的特征参数作为图像的特征向量，用于图像之间的特征比较，L2-范数作为图像特征向量之间的距离相似度度量。

4　实验

4.1　数据集及实验设置

本文在FASHION-MNIST[9]和CIFAR10[10]数据集上进行了图像检索实验。FASHION-MNIST数据集由60000张训练图像和10000张测试图像组成，共分为10个类。每个图像都是28×28的灰度图像，数据集中的部分图像如图2所示。CIFAR10包含60000张32×32的彩色图像，涵盖了各种现实生活图像，训练集包含50000张图像，测试集包含10000张图像，共分为10个类别，数据集中的部分图像如图3所示。本文在keras环境下进行了网络训练以及图像检索实验，采用Adam作为梯度下降算法，Epochs为100，Batch size的大小设置为128，学习率设置为0.001，Routing迭代次数设置为3。

4.2　评价指标

查准率和查全率是常用的图像检索性能评价指标[11]。查准率P描述的是在前N个返回结果中检索到的相关图像的数量除以检索图像的数量，如公式(6)所示：

$$P=\frac{TP}{TP+FP} \tag{6}$$

其中，TP表示检索到的相关图像的数量，FP表示检索到的不相关图像的数量。

查全率R描述的是在前N个返回结果中检索到的相关图像的数量除以图像库中所有相关图像的数量，如公式(7)所示：

$$R=\frac{TP}{TP+FN} \tag{7}$$

其中，TP表示检索到的相关图像的数量，FN表示图像库中未检索到的相关图像的数量。其预测值与实际值之间的对应关系如图4所示。

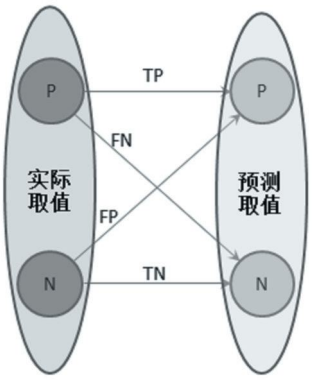


图4　训练集的预测值与实际值之间的关系

由此我们给出一种更为综合的评价指标$Accuracy$，将查准率与查全率相结合，如公式(8)所示。

$$Accuracy=\frac{TP+TN}{TP+FN+FP+TN} \tag{8}$$

4.3　实验及结果分析

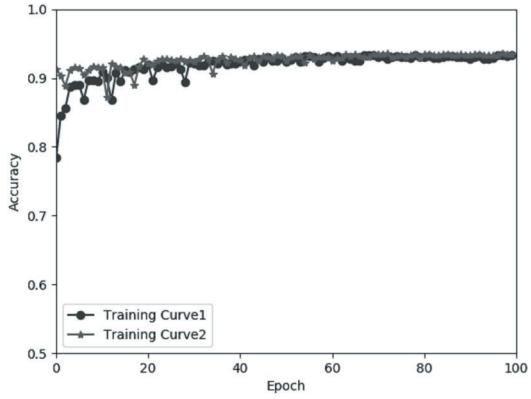
本文首先在FASHION-MNIST数据集上进行了网络模型训练，重构网络采用的是3层二维反卷积网络，各层的参数设置如表1所示。整个训练过程分为两个阶段，两个阶段的区别在于损失函数中参数不同，在第一阶段中$m^+=0.9$，$m^-=0.1$；在第二阶段中$m^+=0.95$，$m^-=0.05$。整个训练过程的分类准确率曲线如图5所示，红色曲线代表第一个阶段的分类准确率，绿色曲线代表第二个阶段的分类准确率，从图中可以看出前30个周期的训练过程中存在波动，且在第二个阶段中增大损失函数中的边界控制值可使训练过程更加稳定并提高分类准确率，测试集的最高分类准确率达到93.56%。

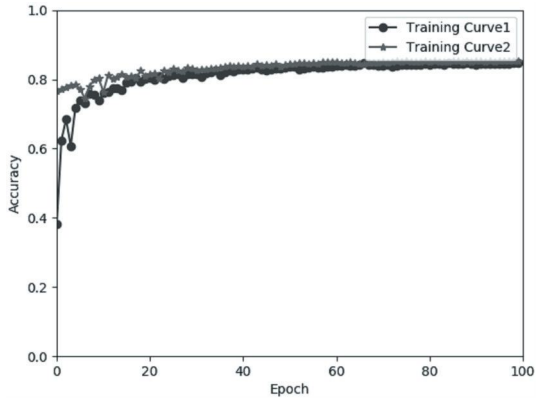
表1　FASHION-MNIST的重构网络参数

卷积核个数	卷积核大小	步长
32	3×3	2
16	3×3	2
1	3×3	1

将训练集作为基本图像库，测试集作为样例图像库，DigitCaps层中的特征参数作为图像的特征向量。训练后可得基本图像库与样例图像库中每一图像的特征向量以及对应的预测类标签，然后，根据预测类标签对基本图像库中的图像进行分类，得到各个类别的子图像库，再将样例图像库中的样例图像划分到某一类别的子图像库中。最后，我们在该类别中查找与样例图像相似的图像，缩小了相似图像的搜索范围。本文尝试将L2-范数作为图像特征向量之间的距离相似度度量，用于图像之间的特征比较。接下来，我们使用查准率和查全率作为衡量检索性能的评价指标，选取前100

张检索到的相似图像作为返回图像，则相应的查准率为92.72%，查全率为1.85%。





接下来在CIFAR10数据集上按照同样的步骤进行了检索实验。考虑到CIFAR10中的图像比FASHION-MNIST中的图像内容更加丰富和复杂，为了更好地获取图像底层的卷积信息，本文首先将CIFAR10中的图像尺寸从32×32×3调整到64×64×3，将64×64×3的图像投入到CapsNet模型中，相应的网络重构部分采用的是4层二维反卷积网络，各层的参数设置如表2所示。整个训练过程的分类准确率曲线如图6所示，红色曲线代表第一个阶段的分类准确率，绿色曲线代表第二个阶段的分类准确率，从图中可以看出在第一阶段的20个训练周期处分类准确率接近80%，能够快速提高模型的训练效果，后期的训练结果逐渐趋于稳定。选取前100张检索到的相似图像作为返回图像，则相应的查准率为84.74%，查全率为1.70%。

表2　CIFAR10的重构网络参数

卷积核个数	卷积核大小	步长
64	3×3	1
32	3×3	2
16	3×3	2
3	3×3	2

5　结论

本文对基于CapsNet的图像检索算法进行了研究。首先，训练神经网络模型来获得图像的特征向量以及预测类标签，将分类与检索任务相结合，找到与样例图像属于同一类的所有图像的集合，然后在该集合中查找样例图像的相似图像。实验结果表明，本文采用的CapsNet模型，可以较好地提取出图像的特征，并且在少量训练周期内可以快速地提高网络的训练准确率。本文仅对简单的灰度图像和彩色图像进行了实验，但CapsNet作为一种新型的网络模型，为我们提供了充分学习图像内在空间关系和仿射变换的方法，在未来使用更精巧的CapsNet架构将会产生更好的结果。

参考文献：

[1] Rajasegaran J，Jayasundara V，Jayasekara S，et al. DeepCaps：Going Deeper With Capsule Networks[C]. arXiv：Computer Vision and Pattern Recognition，2019：10725-10733.

[2] 陈宝盈. 基于深度学习的图像检索技术[D]. 南京：南京邮电大学，2019.

[3] 曹川. 基于深度学习的人脸识别算法研究[D]. 绵阳：西南科技大学，2019.

[4] 贺钰博，刘坤. 基于卷积神经网络的海面显著性目标检测[J/OL]. 计算机工程与应用，2020-02-21.

[5] 李新叶，宋维. 基于深度学习的图像语义分割研究进展[J]. 科学技术与工程，2019，19(33)：21-27.

[6] Ilyas A，Engstrom L，Athalye A，et al. Query-Efficient Black-box Adversarial Examples (superceded)[J]. arXiv：Computer Vision and Pattern Recognition，2017.

[7] Goodfellow I，Shlens J，Szegedy C，et al. Explaining and Harnessing Adversarial Examples[J]. arXiv：Machine Learning，2014.

[8] Sabour S，Frosst N，Hinton G E，et al. Dynamic Routing Between Capsules[J]. arXiv：Computer Vision and Pattern Recognition，2017.

（下转第56页）

[9] Xiao H，Rasul K，Vollgraf R，et al. Fashion-MNIST：a Novel Image Dataset for Benchmarking Machine Learning

Optimization of WSN Coverage Based on Artificial Bee Colony Algorithm

ZHANG Jie SU Qian HAN Zhong-tai
(Shanxi Teachers University, Linfen 041000, Shanxi)

Abstract: The rise of Internet of things technology, wireless communication and computer technology in recent years has attracted the scientific community's attention to wireless sensor networks, and the study of reasonable distribution coverage of detection areas. It is also necessary to maximize coverage. Artificial bee colony algorithm is a kind of optimization method which imitates the bee behavior. It can obtain more superior convergence results for unconstrained numerical optimization problems. Because the existence of artificial bee colony algorithm is easy to be limited to the local optimal solution, the process of the intermediate stagnation problem, the need for a longer search time, an improved artificial bee colony algorithm is proposed, which can speed up the convergence speed in the later stage. The improved artificial bee colony algorithm can effectively reduce the redundancy and prolong the lifetime of the sensor network by optimizing the node coverage.

Keywords: Artificial Bee Colony Algorithm; wireless sensor network; cluster intelligence; coverage optimization

（上接第18页）

Algorithms[J]. arXiv:Learning,2017.

[10] Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. Technical Report TR-2009, University of Toronto, Toronto. 2009.

[11] Chatzichristofis S A, Iakovidou C, Boutalis Y S, et al. Mean Normalized Retrieval Order (MNRO): a new content-based image retrieval performance measure[J]. Multimedia tools and applications, 2014, 70(3): 1767-1798.

A Method for Image Retrieval with Capsule Network

HUANG Jing YANG Shu-guo* LIU Zi-zheng
(Qingdao University of Science and Technology, Qingdao 266061, Shandong)

Abstract: As an emerging network structure, the capsule network uses vector output instead of scalar output, which can capture the spatial relationship between image features and improve the limitations of convolutional neural network. This paper firstly trains the capsule network to achieve image classification, obtains the predictive label of the image, determines the category of the query image, and then uses the feature parameters in the digital capsule layer of the network as the feature vector of the image. The feature vector is used to find images similar to the query image in the category set of the query image. In this paper, experiments are carried out on the FASHION-MNIST and CIFAR10 datasets respectively. The experimental results show that the proposed method can better extract the features of the images and obtain good image retrieval results.

Keywords: image retrieval; capsule network; feature extraction